



Notice d'utilisation du programme d'analyse factorielle en référence à un modèle

Brigitte Escofier

► To cite this version:

Brigitte Escofier. Notice d'utilisation du programme d'analyse factorielle en référence à un modèle.
[Rapport de recherche] RR-0616, INRIA. 1987. inria-00075938

HAL Id: inria-00075938

<https://inria.hal.science/inria-00075938>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Volveau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France

Tél. (1) 39 63 55 11

Rapports de Recherche

N° 616

**NOTICE D'UTILISATION
DU PROGRAMME D'ANALYSE
FACTORIELLE EN RÉFÉRENCE
A UN MODÈLE**

Brigitte ESCOFIER

Février 1987

Campus Universitaire de Beaulieu
Avenue du Général Leclerc
35042 - RENNES CÉDEX
FRANCE
Tél. : (99) 36.20.00
Télex : UNIRISA 95 0473 F

**NOTICE D'UTILISATION DU PROGRAMME D'ANALYSE
FACTORIELLE EN REFERENCE A UN MODELE**

USING GIVE FOR THE COMPUTER PROGRAM OF CORRESPONDENCE ANALYSIS RELATED TO A MODEL

Brigitte **ESCOFIER**

IRISA

Campus de Beaulieu - 35042 RENNES CEDEX

Publication Interne n° 325 - Décembre 86

16 pages

Résumé : L'analyse factorielle des correspondances a été généralisée à l'analyse de la différence entre un tableau de données et un modèle. Dans le cas de l'analyse des correspondances, le modèle est celui de l'indépendance. Ce rapport contient une présentation brève de la méthode et la mise en oeuvre du programme correspondant.

Abstract :

Correspondence analysis has been generalized to the analysis of the distance between a given table of data to any model (For C.A., the model is defined by the hypothesis of independence). This report contains a short presentation of the method and a note of the correspondent program.

NOTICE D'UTILISATION DU PROGRAMME D'ANALYSE FACTORIELLE EN RÉFÉRENCE À UN MODÈLE

1. BUT DU PROGRAMME

Le but de ce programme est l'analyse factorielle d'un tableau K_{IJ} en référence à un tableau modèle de même dimension M_{IJ} .

Une présentation plus complète des aspects théoriques de cette méthode est donnée dans [1] et [2]. Nous nous contentons de rappeler les notions utiles à l'application du programme.

Les principes et les résultats sont tout à fait analogues à ceux de l'analyse des correspondances, dont cette méthode est une généralisation. Deux nuages de points $N(I)$ et $N(J)$ représentent les écarts respectifs des lignes et des colonnes du tableau à celle du modèle. Ces nuages sont projetés sur leurs axes d'inertie. Les moments d'inertie des deux nuages sont égaux et des formules de transition lient les facteurs de même rang. Dans la définition des nuages de points interviennent deux vecteurs P_I et Q_J de dimension respective I et J . Pour chacun de ces vecteurs, il y a 4 options dans le programme. P_I est : soit la marge sur I de K_{IJ} , soit celle de M_{IJ} , soit leur 1/2 somme, soit encore une donnée extérieure.

Les résultats donnés par le programme sont :

- les valeurs propres ou inertie de chaque facteur ainsi que les pourcentages d'inertie extraits,
- les coordonnées des projections des points de $N(I)$ et de $N(J)$ sur leurs axes d'inertie ; les coordonnées des projections des centres de gravité de ces nuages (qui ne sont pas toujours centrés),
- les qualités de représentation et les contributions à l'inertie de chaque point et du centre de gravité des deux nuages,
- les représentations graphiques des facteurs.

Il y a la possibilité de mettre en éléments supplémentaires certaines lignes ou certaines colonnes.

2. NOTATIONS ET FORMULES

Soient K_{IJ} et M_{IJ} deux tableaux de nombres positifs ou nuls de I lignes et J colonnes. Soient P_I et Q_J deux tableaux de nombres positifs de dimension respective I et J .

$$\begin{aligned} K_{IJ} &= \{K_{ij} ; i \in I, j \in J\} & M_{IJ} &= \{M_{ij} ; i \in I, j \in J\} \\ P_I &= \{P_i ; i \in I\} & Q_J &= \{Q_j ; j \in J\} \end{aligned}$$

Notons avec les minuscules correspondantes k_{IJ} , m_{IJ} , p_i et q_j les tableaux obtenus en divisant les éléments des 4 tableaux précédents par leur somme. (calculée sur les lignes et les colonnes actives seulement).

Dans le nuage des lignes, situé dans l'espace R_J , la ligne i est représentée par un point de poids p_i et de coordonnées :

$$\left\{ \frac{k_{ij}}{p_i} - \frac{m_{ij}}{p_i} ; j \in J \right\}$$

Un point i est éloigné de l'origine si la ligne i du tableau k_{IJ} s'écarte beaucoup de celle du modèle m_{IJ} . L'espace R_J est muni de la distance du χ^2 de centre Q_J . La distance entre deux points i et i' est :

$$D^2(i, i') = \sum_j \left\{ \left(\frac{k_{ij}}{p_i} - \frac{m_{ij}}{p_i} \right) - \left(\frac{k_{i'j}}{p_{i'}} - \frac{m_{i'j}}{p_{i'}} \right) \right\}^2 \frac{1}{q_j}$$

Deux points i et i' sont proches si les écarts entre les lignes correspondantes de k_{IJ} et m_{IJ} se ressemblent pour tout j .

Le nuage $N(J)$ est défini symétriquement en inversant I et J .

Les facteurs sont les projections de ces nuages sur leurs axes d'inertie à l'origine, dans l'ordre décroissant de leur moment d'inertie. Les moments d'inertie des deux nuages sont égaux deux à deux. Notons F_s et G_s les facteurs d'ordre s sur I et J et λ_s le moment d'inertie associé. Entre F_s et G_s , on a les relations :

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j \in J} \left(\frac{k_{ij} - m_{ij}}{p_i} \right) G_s(j)$$

$$G_s(j) = \frac{1}{\sqrt{\lambda_s}} \sum_{i \in I} \left(\frac{k_{ij} - m_{ij}}{q_j} \right) F_s(i)$$

Dans la représentation simultanée des nuages $N(I)$ et $N(J)$, les relations entre les positions des éléments des deux nuages peuvent se traduire ainsi : un point i est situé du côté des points j auxquels il s'associe plus dans le tableau que dans le modèle ($k_{ij} > m_{ij}$) et à l'opposé de ceux auxquels il s'associe moins dans le tableau que dans le modèle ($k_{ij} < m_{ij}$). Et réciproquement (pour j relativement aux éléments i).

Le centre de gravité du nuage $N(I)$ a pour coordonnées $k_{.j} - m_{.j}$ où $k_{.j}$ et $m_{.j}$ désignent les marges sur J des tableaux k_{ij} et m_{ij} . Si ces marges ne sont pas égales, le nuage n'est pas centré. De même pour le nuage $N(J)$. La projection des centres de gravité des 2 nuages sur chaque facteur est calculée. On calcule aussi la contribution du centre de gravité à l'inertie des facteurs, qui montre l'influence du décentrage du nuage sur la détermination du facteur. Ce pourcentage doit être assez faible.

Les contributions à l'inertie et les qualités de représentation des éléments des deux nuages se définissent comme en analyse des correspondances.

Connaissant les facteurs F_s et G_s , les inerties λ_s ainsi que les vecteurs p_i et q_j la différence entre le tableau k_{IJ} et le modèle m_{IJ} peut être reconstitué exactement par la formule :

$$k_{ij} - m_{ij} = p_i q_j \sum_s \frac{1}{\sqrt{\lambda_s}} F_s(i) G_s(j)$$

Les résultats coïncident avec ceux de l'analyse des correspondances du tableau k_{IJ} si l'on pose :

$$m_{ij} = k_i k_j$$

$$p_i = k_i$$

$$q_j = k_j$$

3. APPLICATIONS

3.1. Tableaux d'échanges ou de trafic

Ce sont des tableaux carrés dont le traitement par l'Analyse des Correspondances pose des problèmes à cause de leur diagonale.

On pose :

$$K_{ii} = 0$$

$$M_{ij} = K_i K_j \quad \text{si } i \neq j$$

$$M_{ii} = 0$$

$$P_i = K_i$$

$$Q_j = K_j$$

3.2. Analyse d'un sous tableau et analyse à marge modifiée

Soit K_{IJ} un sous tableau contenu dans $K_{I',J}$ avec $I \in I'$. Pour faire l'analyse du sous tableau en gardant la métrique définie sur R_J par le tableau entier ainsi que les poids des éléments j , on pose :

$$M_{ij} = K_i K'_j \quad \text{où } K'_j \text{ est la marge du tableau entier}$$

$$P_i = K_i$$

$$Q_j = K'_j$$

Dans les formules ci-dessus, on peut prendre pour K'_j des valeurs quelconques, ce qui permet de faire une analyse des correspondances de K_{IJ} en remplaçant la marge K_J par des valeurs données.

3.3. Analyse des résidus d'un modèle

M_{IJ} peut être un modèle construit à partir de variables extérieures ou de certaines hypothèses.

4. LE PROGRAMME

Il a été écrit à partir du programme d'analyse des correspondances figurant dans la bibliothèque d'analyse des données de l'ADDAD. Une analyse du programme figure dans [3].

4.1. Les données

Elles sont lues sur le fichier de numéro 01. Pour chaque ligne i , on lit successivement le nom de cette ligne, les valeurs des éléments de K_{IJ} suivies des valeurs des éléments de M_{IJ} . S'il y a des lignes supplémentaires, elles doivent figurer en fin de fichier.

4.2. Les paramètres sont lus sur le fichier de numéro 05.

Carte 1 : Titre de l'analyse Format 20 A4

Carte 2 : Paramètres généraux Format V

NI	Nombre de lignes (ensemble I)
NJ	Nombre de colonnes (ensemble J)
NF	Nombre de facteurs calculés
NI2	Nombre de lignes supplémentaires (en fin de tableau)
NJ2	Nombre de colonnes supplémentaires
IOUT=1	Sortie des tableaux (données et modèle)
IF=1	Sortie des facteurs sur I
JF=1	Sortie des facteurs sur J
IGR=1	Des graphiques sont demandés (carte 6 nécessaire)
ISTOCK=1	Stockage des facteurs sur I sur file 25
=2	Stockage des facteurs sur les lignes supplémentaires sur file 07
JSTOCK=1	Stockage des facteurs sur J sur file 24
=2	Stockage des facteurs sur les colonnes supplémentaires sur file 07
JOPT=1	Lecture des $Q(j)$
=2	$Q(j) = k_j$
=3	$Q(j) = (k_j + m_j)/2$
=4	$Q(j) = m_j$
IOPT=1	Lecture des $P(i)$
=2	$P(i) = k_i$
=3	$P(i) = (k_i + m_i)/2$
=4	$P(i) = m_i$

Carte 3	Noms des variables (Format 20 A4)
Carte 4	Position des colonnes supplémentaires (Format 80I1) (si NJ2#0)
Carte 5	Format de lecture des données (Format 20 A4)
Carte 6	Demande de graphiques (si IGR#0)(Format 80I1) Pour chaque graphique (8 au maximum) une zone de 10 colonnes
Carte 7	Format de lecture des Q(j) (si JOPT=1)
Carte 8	Lecture des Q(j) (si JOPT=1)
Carte 9	Format de lecture des P(i) (si IOPT=1)
Carte 10	Lecture des P(i) (si IOPT=1)

5. EXEMPLE D'UTILISATION

Le tableau K_{IJ} est une matrice d'échange traitée suivant le procédé indiqué en 3.1.

5.1. Fichier de paramètres

```
essai
  15  15  7  0  0  1  1  1  1  0  0  2  2
  +1  +2  +3  +4  +5  +6  +7  +8  +9  +10  +11  +12  +13  +14  +15
(a2,1x,15f6.0,15f6.0)
2133 4221 1333 4221
```

5.2. Fichier de données

1	0.	259.	45.	14.	204.	1824.	57.	250.	70.	76.	16.	36.	0.	403.	189.	0.	149.	285.	139.	429.	419.	266.
301.	171.	297.	113.	166.	191.	273.	155.															
2	270.	0.	1113.	257.	2483.	1450.	530.	708.	166.	878.	166.	205.	281.	457.	174.	259.	0.	754.	367.	1136.	1109.	704.
797.	454.	787.	300.	440.	505.	723.	410.															
3	34.	525.	0.	1001.	1493.	32.	143.	62.	133.	207.	327.	549.	226.	133.	0.	139.	212.	0.	198.	612.	598.	379.
430.	244.	424.	162.	237.	272.	390.	221.															
4	0.	116.	1389.	0.	425.	100.	99.	220.	27.	111.	215.	1037.	26.	152.	117.	114.	173.	332.	0.	500.	488.	310.
351.	200.	347.	132.	194.	223.	318.	181.															
5	186.	667.	394.	291.	0.	1009.	1577.	148.	123.	1021.	154.	265.	860.	314.	90.	215.	327.	626.	305.	0.	921.	584.
652.	377.	654.	249.	365.	420.	601.	341.															
6	713.	253.	134.	75.	632.	0.	595.	1675.	563.	250.	29.	0.	118.	507.	297.	166.	252.	482.	235.	727.	0.	450.
510.	290.	504.	192.	282.	323.	463.	262.															
7	0.	181.	78.	41.	763.	148.	0.	24.	396.	964.	104.	38.	745.	25.	87.	102.	155.	296.	144.	447.	436.	0.
313.	178.	310.	118.	173.	199.	284.	161.															
8	51.	31.	68.	0.	133.	1094.	109.	0.	107.	92.	0.	28.	39.	1831.	491.	117.	177.	340.	166.	513.	501.	318.
9	0.	205.	355.	135.	199.	228.	326.	185.														
165.	0.	163.	62.	91.	105.	150.	85.															
10	14.	103.	492.	177.	353.	104.	528.	209.	568.	0.	315.	408.	551.	191.	130.	117.	179.	342.	167.	516.	503.	319.
362.	206.	0.	136.	200.	229.	328.	186.															
11	0.	21.	160.	83.	81.	33.	23.	20.	64.	248.	0.	110.	106.	21.	0.	27.	42.	80.	39.	121.	118.	75.
95.	48.	84.	0.	47.	54.	77.	44.															
12	0.	53.	310.	260.	156.	0.	0.	0.	0.	82.	481.	0.	131.	0.	0.	42.	63.	122.	59.	183.	179.	113.
123.	73.	127.	48.	0.	81.	117.	66.															
13	0.	66.	21.	0.	151.	40.	421.	24.	43.	248.	26.	0.	0.	25.	0.	30.	46.	88.	43.	132.	129.	82.
93.	53.	92.	35.	51.	0.	34.	48.															
14	327.	43.	0.	63.	206.	301.	42.	1362.	0.	40.	54.	90.	35.	0.	774.	109.	165.	317.	154.	477.	466.	295.
335.	190.	330.	126.	185.	212.	0.	172.															
15	0.	0.	0.	26.	26.	20.	28.	159.	591.	102.	0.	0.	0.	403.	0.	38.	58.	112.	54.	168.	164.	104.
118.	67.	117.	45.	65.	75.	107.	0.															

5.3. Résultats

COMPARAISON DE 2 TABLEAUX DE FREQUENCE essai

NI	NJ	NF	NI2	NJ2	IOU	IF	JF	IGR	ISTOCK	JSTOCK	JOPT	IOPT
15	15	7	0	0	1	1	1	1	0	0	2	2
+1(1)			+2(2)			+3(3)			+4(4)	+5(5)		+6(6)
+11(11)			+12(12)			+13(13)			+14(14)	+15(15)		

(a2,1x,15f6.0,15f6.0)

tableau
des
données

NOMJ (J) :	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15		
PJ (J)	1626.	2472.	4738.	2306.	7140.	6971.	4423.	5009.	2851.	4947.	1887.	2766.	3177.	4545.	2577.	57435.	
1	!	0.	269.	45.	14.	204.	1824.	57.	250.	70.	76.	16.	36.	0.	403.	189.	3453.
2	!	270.	0.	1113.	257.	2483.	1450.	530.	708.	166.	878.	166.	205.	281.	457.	174.	9138.
3	!	34.	585.	0.	1001.	1493.	32.	143.	62.	133.	207.	327.	549.	226.	133.	0.	4925.
4	!	0.	106.	1337.	0.	425.	100.	99.	220.	27.	111.	215.	1037.	26.	152.	117.	4024.
5	!	186.	667.	894.	281.	0.	1009.	1577.	148.	123.	1021.	154.	265.	860.	314.	90.	7589.
6	!	713.	258.	134.	75.	632.	0.	595.	1675.	563.	250.	29.	0.	118.	507.	297.	5846.
7	!	0.	181.	78.	41.	763.	148.	0.	24.	396.	964.	104.	38.	745.	25.	87.	3594.
8	!	51.	81.	68.	0.	133.	1094.	109.	0.	107.	92.	0.	28.	39.	1831.	491.	4124.
9	!	31.	34.	34.	28.	34.	316.	271.	148.	0.	628.	0.	0.	59.	83.	228.	1894.
10	!	14.	103.	492.	177.	353.	104.	528.	209.	568.	0.	315.	408.	551.	191.	130.	4148.
11	!	0.	21.	160.	83.	81.	33.	23.	20.	64.	248.	0.	110.	106.	21.	0.	970.
12	!	7.	53.	317.	260.	156.	0.	0.	0.	0.	82.	481.	0.	131.	0.	0.	1473.
13	!	0.	66.	21.	0.	151.	40.	421.	24.	43.	248.	26.	0.	0.	25.	0.	1065.
14	!	327.	43.	0.	63.	206.	801.	42.	1362.	0.	40.	54.	90.	35.	0.	774.	3837.
15	!	0.	0.	0.	26.	26.	20.	28.	159.	591.	102.	0.	0.	0.	403.	0.	1355.

modèle

NOMJ(J) :	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15		
PJ(J)	1529.	2080.	4332.	2146.	6196.	6261.	4145.	4650.	2756.	4591.	1853.	2695.	3117.	4241.	2517.	53109.	
1	!	0.	149.	285.	139.	429.	419.	266.	301.	171.	297.	113.	166.	191.	273.	155.	3354.
2	!	259.	0.	754.	367.	1136.	1109.	704.	797.	454.	787.	300.	440.	505.	723.	410.	8745.
3	!	139.	212.	0.	198.	612.	598.	379.	430.	244.	424.	162.	237.	272.	390.	221.	4518.
4	!	114.	173.	332.	0.	500.	488.	310.	351.	200.	347.	132.	194.	223.	318.	181.	3863.
5	!	215.	327.	626.	305.	0.	921.	584.	662.	377.	654.	249.	365.	420.	601.	341.	6647.
6	!	166.	252.	492.	235.	727.	0.	450.	510.	290.	504.	192.	282.	323.	463.	262.	5138.
7	!	102.	155.	296.	144.	447.	436.	0.	313.	178.	310.	118.	173.	199.	284.	161.	3316.
8	!	117.	177.	340.	166.	513.	501.	318.	0.	205.	355.	135.	199.	228.	326.	185.	3765.
9	!	54.	92.	156.	76.	235.	230.	146.	165.	0.	163.	62.	91.	105.	150.	85.	1800.
10	!	117.	179.	342.	167.	516.	503.	319.	362.	206.	0.	136.	200.	229.	328.	186.	3790.
11	!	27.	42.	80.	39.	121.	118.	75.	85.	48.	84.	0.	47.	54.	77.	44.	941.
12	!	42.	63.	122.	59.	133.	179.	113.	123.	73.	127.	48.	0.	81.	117.	66.	1401.
13	!	30.	46.	88.	43.	132.	129.	82.	93.	53.	92.	35.	51.	0.	84.	48.	1006.
14	!	109.	165.	317.	154.	477.	466.	295.	335.	190.	330.	126.	185.	212.	0.	172.	3533.
15	!	38.	58.	112.	54.	168.	164.	104.	113.	67.	117.	45.	65.	75.	107.	0.	1292.

DIFFERENCE ENTRE TABLEAU ET MODELE (en millième)

NOMJ(J) :	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15	G(I)
G(J)	23	43	82	40	124	121	77	87	49	86	32	48	55	79	44	
1	0	1	-4	-2	-4	23	-4	-1	-2	-4	-1	-2	-3	1	0	60
2	0	0	5	-2	21	4	-4	-2	-5	0	-2	-4	-4	-5	-4	159
3	-2	6	0	13	14	-10	-4	-7	-2	-4	2	5	-1	-5	-4	85
4	-2	-1	17	0	-2	-7	-4	-2	-3	-4	1	14	-3	-3	-1	70
5	0	5	3	0	0	0	16	-9	-4	5	-2	-2	7	-5	-4	132
6	9	0	-6	-3	-2	0	1	19	4	-5	-3	-5	-4	0	0	101
7	-1	0	-4	-1	4	-5	0	-5	3	10	0	-2	9	-4	-1	62
8	-1	-1	-5	-3	-7	2	-4	0	-1	-5	-2	-3	-3	25	5	71
9	0	0	-2	0	-3	1	1	0	0	7	-1	-1	0	-1	2	32
10	-1	-1	2	0	-3	-7	3	-3	6	0	2	3	5	-2	-1	72
11	0	0	1	0	0	-1	-1	-1	0	2	0	1	0	-1	0	16
12	0	0	3	3	0	-3	-2	-2	-1	0	7	0	0	-2	-1	25
13	0	0	-1	0	0	-1	5	-1	0	2	0	0	0	-1	0	18
14	3	-2	-5	-1	-5	5	-4	17	-3	-5	-1	-1	-3	0	10	66
15	0	-1	-2	0	-2	-2	-1	0	9	0	0	-1	-1	5	0	23

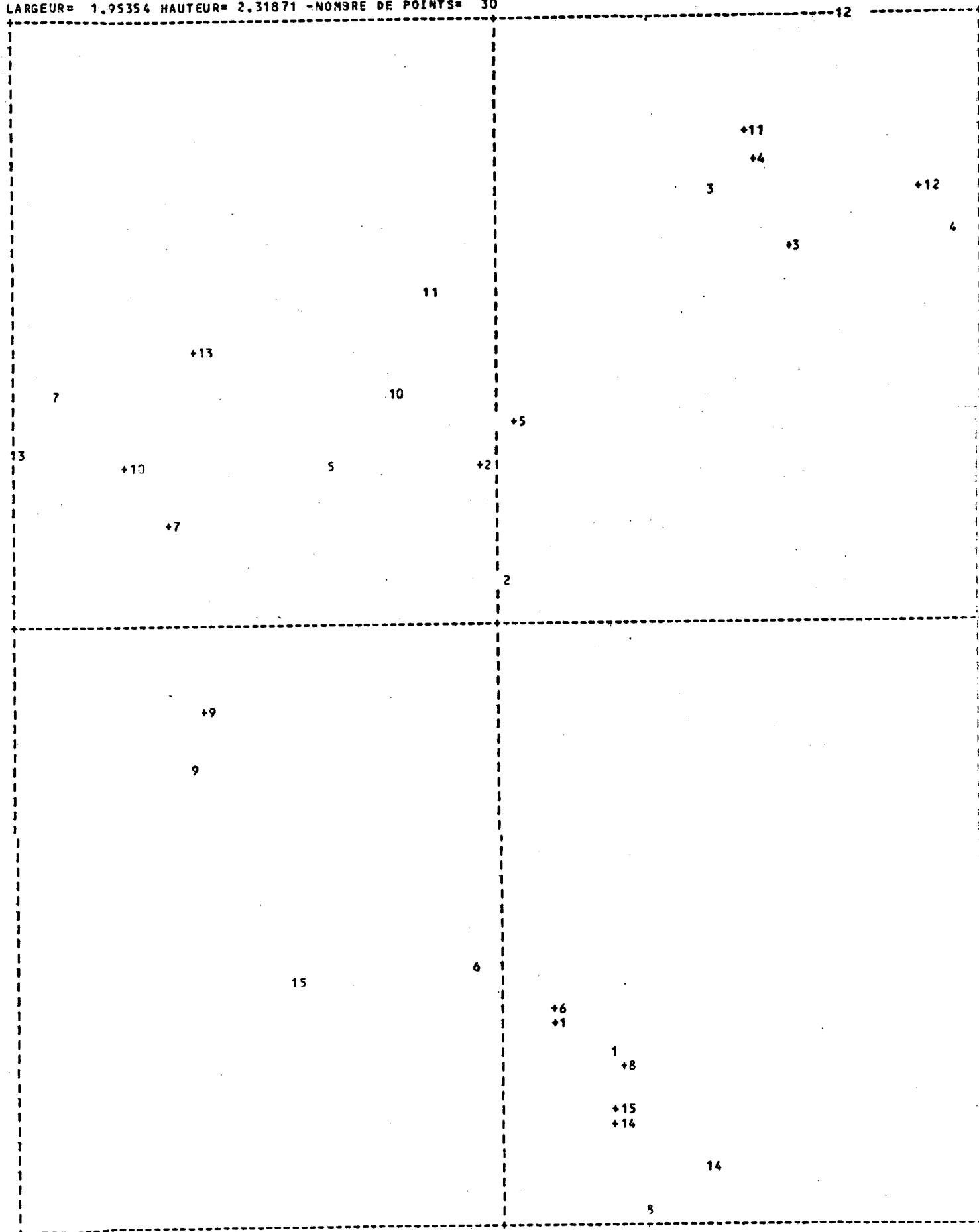
LES VALEURS PROPRES

VAL(1)=0.44907634

INUM	ITER	VAL PROPRE	POURCENTI	CUMUL	I*	HISTOGRAMME DES VALEURS PROPRES DE LA MATRICE
1	1	0.44907634	34.731	34.731	1*	*****
2	1	0.21633573	16.731	51.462	1*	*****
3	1	0.15075323	11.659	63.121	1*	*****
4	2	0.14372012	11.115	74.236	1*	*****
5	1	0.10144633	7.846	82.082	1*	*****
6	2	0.07021035	5.430	87.512	1*	*****
7	2	0.05323372	3.895	91.397	1*	*****
8	2	0.04603493	3.560	94.957	1*	*****
9	2	0.03061644	2.368	97.325	1*	*****
10	2	0.02088659	1.615	98.940	1*	*****
11	2	0.00952436	0.737	99.677	1*	*****
12	1	0.00281027	0.217	99.894	1*	*****
13	5	0.00096073	0.074	99.967	1*	*****
14	4	0.00040007	0.031	100.000	1*	*****
15	1	0.00000597	0.000	100.000	1*	*****

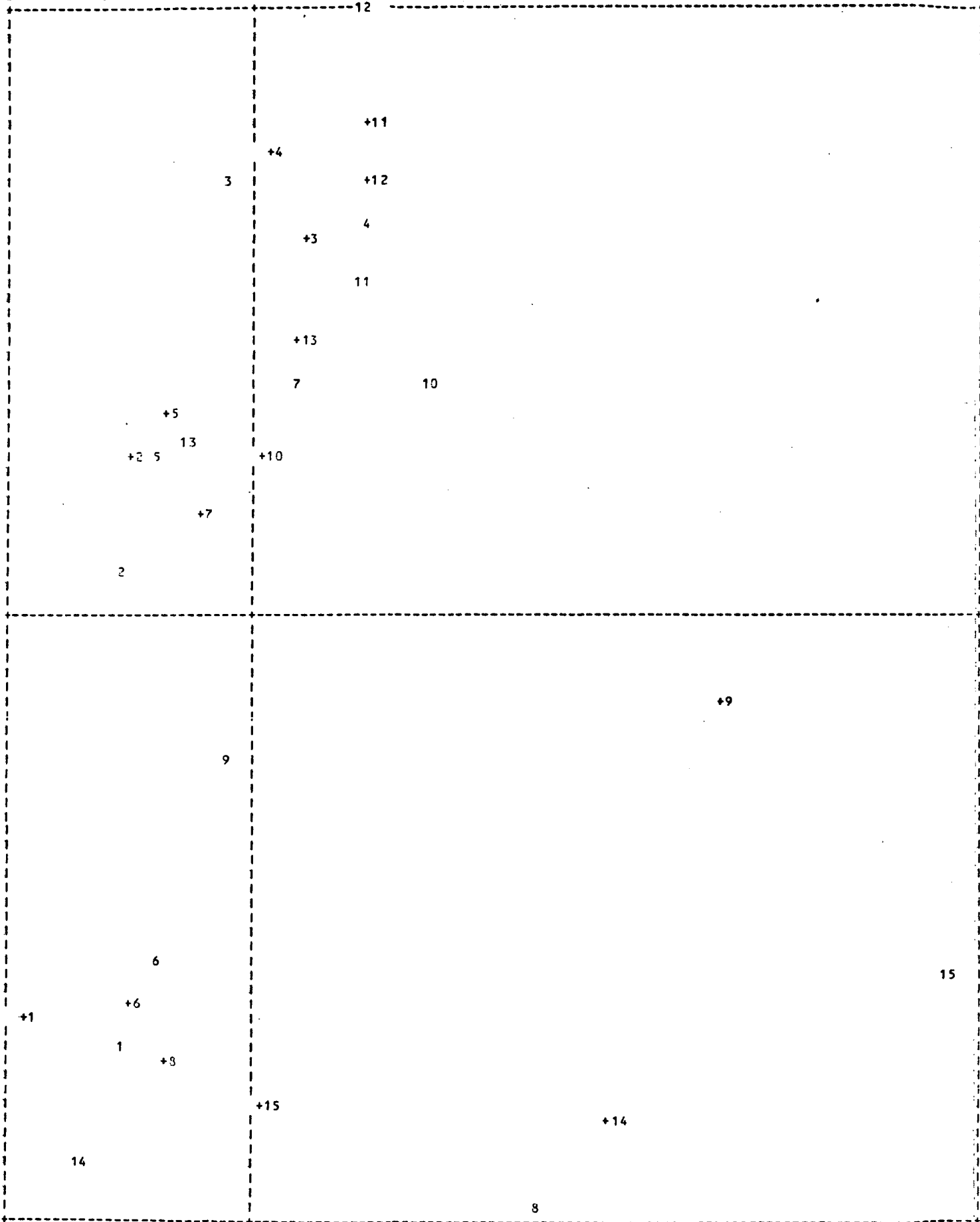
AXE HORIZONTAL(2)--AXE VERTICAL(1)--TITRE:essai

LARGEUR= 1.95354 HAUTEUR= 2.31871 -NOMBRE DE POINTS= 30



AXE HORIZONTAL(3)--AXE VERTICAL(1)--TITRE:essai

LARGEUR= 2.22157 HAUTEUR= 2.31871 -NOMBRE DE POINTS= 30



BIBLIOGRAPHIE

- B. ESCOFIER [1] Analyse Factorielle en référence à un modèle
Application à l'analyse de tableaux d'échange.
Rapport IRISA n° 229 - Juillet 1984.
Revue de Statistique Appliquée 1985.
- B. ESCOFIER [2] Analyse de la différence entre deux mesures sur
le produit de deux mêmes ensembles.
Cahiers de l'Analyse des Données - 1983 - n°3.
- P. PUERTOLAS [3] Mise au point d'un programme analysant les écarts
d'un tableau à un modèle.
Rapport de stage IRISA.

